Democracy and Economic Development in Turkey: 
An Exploratory Spatial Data Analysis

Fatih ÇELEBİOĞLU  
Dr., University of Dumlupinar, Faculty of Economics and Administrative Sciences,  
Department of Economics, Kutahya, Turkey  
fcelebi@dumlupinar.edu.tr  

Hüseyin ALTAY  
Dr., University of Bilecik, Faculty of Economics and Administrative Sciences,  
Department of Economics, Bilecik, Turkey  
haltay16@hotmail.com

Abstract: The aim of this paper is to perform an exploratory spatial data analysis on the democracy and development level of the 76 Turkish regions over 1995-2001. While our choropleth maps indicate that the Western part of the country is significantly more developed than the East, the tools of spatial statistics reveal the presence of spatial dependence across provinces. The presence of heterogeneity is reflected in the distribution of LISA statistics. Overall, our results shed new light on the distribution of growth across Turkish regions and its relation with participation rates in general elections.

Key words: Turkey, Democracy, Regional Disparities, Spatial Statistics

Introduction

Democracy is a form of government in which state-power is held by the majority of citizens within a country. All the people should be able to have their say in one way another in everything that affects their lives. It is interesting that almost each developed countries have high level democracy and civil freedom. Furthermore, democracy is related to the level of economic development of a particular country. For this reason, relations between democracy and economic development are intensively examined by economists in the recent years.

The first cross-national study on world democratizations emphasized the conditioning effect of industrialization and economic development was performed by Lipset (1959) in his seminal study. Bollen (1979) explores the relationship between development timing and political democracy. Arat (1988) explore democratic instability and economic development. Burkhart and Lewis-Beck (1994) finds that economic development "causes" democracy, but democracy does not "cause" economic development. Muller (1995) investigates that the relationship between the level of economic development and the level of democracy found in most quantitative cross-national research.


In this study, we examine relations between democracy and economic development in terms of regional (provincial) level in Turkey by using Exploratory Spatial Data Analysis (ESDA).

Turkey and Regional Development

Turkey is formally composed of several provinces used as administrative units. The definition of regions is only used for geographic classification purposes (for example Marmara, Aegean, Southeastern areas) and to cluster provinces according to their level of economic development. For instance, the provinces located in the Southeastern and Eastern Anatolia areas are known to be lagging behind in economic and social terms.
A couple of reasons have been highlighted in the past to justify the East-West divide that has marked the Turkish regional economies for a couple of decades (Ates et al. 2000; Balkir 1995; Gezici and Hewings 2004). They are, among others, inequalities in salaries (Elveren and Galbright, 2008), the dependence on agriculture and weakness of industrial sector (Ozaslan et al. 2006), the divide in the education level (Ozturk 2002), the migratory flows from the east to the west (Kirdar and Saracoglu 2007), and the lack of private investment in the east (Deliktas et al. 2008). However, it is very difficult to assess the extent to which the phenomena above are the reason or the consequence for the divide observed within Turkey.

Following the spirit of the literature cited above, the aim of this paper is to investigate inequalities across the 76 Turkish regions over 1995-2001 by means of an exploratory spatial data analysis (ESDA). It is a set of techniques used to describe and visualize spatial distributions, identify atypical locations or spatial outliers, discover patterns of spatial association, clusters or hot spots, and suggest spatial regimes or other forms of spatial heterogeneity (Anselin 1988 and 1999). Several ESDA have been performed on the issue of regional inequalities. For instance, Dall’erba (2005), Ezcurra et al. (2007), Battisti and Di Vaio (2008) focus on the EU regions.

ESDA offers the opportunity to compare the differences between the eastern and western provinces by means of choropleth maps, box plots and scatter plots and measure the extent of spatial autocorrelation.

**Data analysis**

Our dataset comes from the Turkish Statistical Institute and the State Planning Organization. They represent for each region the level of per capita income in 1995, the growth rate of per capita income over 1995-2001, participation rates to general election of the region in 1995 and province level literacy rates in 1995. All data are expressed in 1987 constant prices. The time frame we use (1995-2001) is limited by data availability. Indeed, data before and after that period simply do not exist at the regional level. As a result, even if Turkey currently counts 81 provinces, we are obliged to work with the 76 provinces that correspond to that period.

**Mapping the Distributions**

We start our analysis with the quartile maps of the distribution of our variables for each province. Figure 1 displays the distribution of the regional growth rate of per capita GDP relative to Turkey’s average over 1995-2001. The darker areas indicate a greater level of relative growth. It appears from this map that the distribution of growth is pretty random, which is an idea that will need to be assessed in the next section.

Figure 2 displays the distribution of regional per capita GDP levels in 1995 relative to Turkey’s average. A clear core-periphery (or east / west) pattern appears in this map, with the core composed of the richest regions, whereas the peripheral regions are also the poorest ones. This confirms the findings of the various studies mentioned in the introduction above. In the Western part of the country, the coastal areas and the province of the capital city are clearly better off than the rest of the country. This is because trade, industry and tourism are developed in these areas.
Figure 3 shows participation rates to 1995 general elections. This result is clearly indicates that participation rates in Turkey’s East part are lower than West part.

Figure 4 may give us more insights into the East-West disparities mentioned so far. Indeed, as can be seen on this quartile map, literacy rates as % of population in Turkey (in 1995 and relative to Turkey’s average) is much greater in the West than in the East.

As a result, it can be linkage that low level participation to democratic elections and low literacy rates in the East part of Turkey.
**Box Plots**

The box plot is another tool of ESDA. Designed by John Tukey (1977), box plots display five interesting pieces of information about a dataset: the lowest value, the lower quartile of the distribution (25% of the cumulative distribution, noted Q1), the median (Q2), the upper quartile (75% of the cumulative distribution, noted Q3), and the highest value. The median is represented by the line in the center of the rectangular box. In addition, a box plot displays the outliers which are defined as the values above or below a given multiple (either 1.5 or 3) of the difference between the first and third quartile. For instance, a lower outlier corresponds to a value below \([Q1-1.5\times(Q3-Q1)]\) and an upper outlier is defined as a value above \([Q3+1.5\times(Q3-Q1)]\). The thin line on the upper part of box plots is called the hinge, here corresponding to the default criteria of 1.5 times the difference between the first and third quartile (Thompson 2003).

The box plots of our variables appear in figures 5 to 8. They show that Bolu and Zonguldak are the only (upper) outliers in the distribution of provincial growth rate; while only Kocaeli is the province with the highest value of per capita GDP in 1995 but it is not an outlier. Manisa is the province with the highest value of participation rates to general elections and Istanbul has the highest value of literacy rates but they are not an outlier also.

Canakkale is the province with the lowest value of growth rate in the period of 1995-2001. There are two lowest values ( Ağrı and Mus) in the distribution of log of per capita GDP in 1995. While Ağrı, Batman, Bitlis, Diyarbakır, Hakkari, Mardin, Muş, Şanlı URfa, Siirt and Van are the provinces with the lowest values in the distribution of literacy rates, other some provinces ( Ağrı, Ardahan, Bingöl, Bitlis, Diyarbakır, Erzincan, Gümüşhane, Iğdır, Kars, Rize, Siirt, Şırnak and Tunceli) have the lowest values in the distribution of the share of the population with a university degree.

 Quartile maps and box plots are useful tools to get some insights into the distribution of a variable. However, they do not formally test whether the spatial distribution of a variable is random or not. For instance, the distribution of the per capita income and province level literacy rates across Turkish provinces is marked by two distinct clusters (East vs. West) as can be seen from figures 2 and 4 above. This observation needs to be tested by the formal tools of Exploratory Spatial Data Analysis. It starts with the definition of a spatial weight matrix and continues with the measurement of spatial autocorrelation.

Fig.5 Growth rate for period of 1995-2001 in Turkey

Fig.6 Log of Per Capita GDP in Turkey (1995)
Exploratory Spatial Data Analysis (ESDA)

Spatial Weight Matrix

A spatial weight matrix is the necessary tool to impose a neighborhood structure on a spatial dataset. As usual in the spatial statistics literature, neighbors are defined by a binary relationship (0 for non-neighbors, 1 for neighbors). All our work is performed under GeoDa. We have used a basic approach for defining neighborhood: contiguity (shared borders). Contiguity-based weights matrices include rook and queen. Areas are neighbors under the rook criterion if they share a common border, not vertices. Based on this concept, we decided to create a weight matrix to investigate the distribution of our variables of interest: k nearest neighbor matrix. We present the k nearest neighbor matrix only below:

\[
\begin{align*}
    w_{ij}(k) &= \begin{cases} 
    0 & \text{if } i = j \\
    1 & \text{if } d_{ij} \leq D_i(k) \text{ and } w_{ij}(k) = w_{ij}(k) / \sum_j w_{ij}(k) \text{ for } k = 7 \\
    0 & \text{if } d_{ij} > D_i(k)
    \end{cases}
\end{align*}
\]

where \(d_{ij}\) is great circle distance between centroids of region \(i\) and \(j\) and \(D_i(k)\) is the \(7^{th}\) order smallest distance between regions \(i\) and \(j\) such that each region \(i\) has exactly 7 neighbors. Now that the weight matrix has been defined, we estimate a couple of spatial statistics that will shed some light on the spatial distribution of our variables. The most common of them is Moran’s I which is a measure of global spatial autocorrelation (Anselin 1988).

Moran’s I for Global Spatial Autocorrelation

Spatial autocorrelation refers to the correlation of a variable with itself in space. It can be positive (when high values correlate with high neighboring values or when values correlate with low neighboring values low) or negative (spatial outliers for high-low or low-high values). Note that positive spatial autocorrelation can be associated with a small negative value (e.g., -0.01) since the mean in finite samples is not centered on 1. Spatial autocorrelation analysis includes tests and visualization of both global (test for clustering) and local (test for clusters) Moran’s I statistic (Anselin et al. 2006).
Global spatial autocorrelation is a measure of overall clustering and it is measured here by Moran’s I. It captures the extent of overall clustering that exists in a dataset. It is assessed by means of a test of a null hypothesis of random location. Rejection of this null hypothesis suggests a spatial pattern or spatial structure, which provides more insights about a data distribution that what a quartile map or box plot does. For each variable, it measures the degree of linear association between its value at one location and the spatially weighted average of neighboring values (Anselin et al. 2007; Anselin 1995) and is formalized as follows:

\[
I_I = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(k)x_{ij}x_{jt}}{\sum_{i=1}^{n} \sum_{j=1}^{n} x_{it}x_{jt}}
\]

(2)

Where \( w_{ij} \) is the (row-standardized) degree of connection between the spatial units \( i \) and \( j \) and \( x_{ij} \) is the variable of interest in region \( i \) at year \( t \) (measured as a deviation from the mean value for that year). Values of \( I \) larger (smaller) than the expected value \( E(I) = -1/(n-1) \) indicate positive (negative) spatial autocorrelation. In our study, this value is (-0.0133). There are different ways to draw inference here. The approach we use is a permutation approach with 999 permutations. It means that 999 re-sampled datasets were automatically created for which the \( I \) statistics are computed. The value obtained for the actual dataset has then been compared to the empirical distribution obtained from these re-sampled datasets.

The results of Moran’s I are presented in Table 1 below. All the results indicate a positive spatial autocorrelation, i.e., the value of a variable in one location depends positively on the value of the same variable in neighboring locations. For instance, when the per capita income in one province increases by 1%, the one of its neighbors increases by slightly more than 0.6%. Three out of our four variables of interest are significant (at 1%) with the k_7 nearest neighbor matrix. For this reason, this is the weight matrix we will use in the rest of our study.

**Table 1: Moran’s I and P-Value**

<table>
<thead>
<tr>
<th>Variables</th>
<th>K_7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Growth Rate (1995-2001)</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td>(0.129)</td>
</tr>
<tr>
<td>Log of Per Capita GDP (1995)</td>
<td>0.647</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
</tr>
<tr>
<td>Participation Rates to Gen. Elections (1995)</td>
<td>0.706</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
</tr>
<tr>
<td>Literacy Rates (1995)</td>
<td>0.799</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
</tr>
</tbody>
</table>

Note: p-values are into brackets

Moran’s Scatter plot for Global and Local Spatial Autocorrelation

The Moran scatter plot often complements Moran’s I because it provides an easy way to categorize the nature of spatial autocorrelation into four types: low-low (noted LL), low-high (LH), high-low (HL) and high-high (HH) with the first part refereeing to the studied location itself and the second part refereeing to the neighboring ones. For instance, HH means a high value in the studied area and a high value in the neighboring areas. The four types are reflected in the four quadrants that compose a Moran’s scatter plot. The link between a scatter plot and Moran’s I is reflected by a line of which slope is the value of Moran’s I statistic.

Regions located in quadrants I and III refer to positive spatial autocorrelation, the spatial clustering of similar values, whereas quadrants II and IV represent negative spatial autocorrelation, the spatial clustering of dissimilar values.

Figures 9 to 12 below display the Moran scatter plots of our variables of interest. For both the per capita income, participation rates and literacy rates, positive spatial autocorrelation is reflected by the value of Moran’s I and the fact that most of the provinces are located in quadrants HH and LL with HH displaying a cluster of Western provinces while LL shows a cluster of Eastern provinces. Once again, it reflects the dualistic structure of Turkey’s provinces.
Fig. 9 Growth rate for period of 1995-2001 in Turkey

Fig. 10 Log of Per Capita GDP in Turkey (1995)

Fig. 11 Participation Rates to General elections (1995)

Fig. 12 Province Level Literacy Rates in Turkey (1995)
Table 2 indicates the name of the regions according to their distribution in the Moran scatterplot quadrants. Positive spatial autocorrelation is reflected by the fact that most provinces are in the high-high and low-low quadrants. More precisely, for the per capita GDP, participation rate and literacy rate variables, the average of the Local Moran statistics is proportional to the Global Moran's I value (Anselin 1995; Anselin et al. 2007).

### Table 2 Distribution of Spatial Autocorrelation

<table>
<thead>
<tr>
<th>Growth rate for period of 1995-2001</th>
<th>HH</th>
<th>LL</th>
<th>LH</th>
<th>HL</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Log of province level per capita GDP (1995)</th>
<th>HH</th>
<th>LL</th>
<th>LH</th>
<th>HL</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Participation to General Elections (1995)</th>
<th>HH</th>
<th>LL</th>
<th>LH</th>
<th>HL</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Literacy Rates (1995)</th>
<th>HH</th>
<th>LL</th>
<th>LH</th>
<th>HL</th>
</tr>
</thead>
</table>

### LISA Statistics for Local Spatial Autocorrelation

LISA statistics (Local Indicators of Spatial Autocorrelation) measure, by definition, the presence of spatial autocorrelation for each of the location of our sample. It captures the presence or absence of significant spatial clusters or outliers for each location. Combined with the classification into four types defined in the Moran scatter plot above, LISA indicates significant local clusters (high–high or low–low) or local spatial outliers (high–low or low–high). The average of the Local Moran statistics is proportional to the Global Moran’s I value (Anselin 1995; Anselin et al. 2007).

Anselin (1995) formulated the local Moran’s statistics for each region $i$ and year $t$ as the follows:

$$I_i = \left(\frac{x}{m_0}\right) \sum_j w_{ij} x_j \quad \text{with} \quad m_0 = \sum_i x_i^2 / n \quad (3)$$
where \( w_{ij} \) is the elements of the row-standardized weights matrix \( W \) and \( x_i(x_j) \) is the observation in region \( i \) \( (j) \).

Once again, this result reflects the will of the authorities to counterbalance poverty in the East. We provide the LISA maps (figures 13 to 16) as a visual representation of these results.

**Fig. 13** Cluster Map (Growth Rate 1995-2001)

**Fig. 14** Cluster Map (Log of Per Capita GDP 1995)

**Fig. 15** Cluster Map (Participation Rates in 1995)

**Fig. 16** Cluster Map (Literacy Rates in 1995)
Conclusions

The aim of this paper has been to perform an exploratory analysis of the economic disparities across 76 Turkish provinces. We have investigated the spatial distribution of growth over 1995-2001, of the per capita GDP, participation rates and literacy rates in 1995 across these provinces. First, our quartile maps have revealed the gap between East and West when it comes to per capita GDP, participation rates and literacy rates. Second, the Box plots showed that West Anatolia and the coastal area provinces are upper outliers in the distribution of almost all our variables. When we measure spatial autocorrelation by means of Moran’s I, our results indicate positive (and significant) global autocorrelation for all our variables except growth, thus indicating the geographical location of a province influences its level of income, participation rates and literacy rates.

These results are corroborated by the corresponding Moran’s Scatterplots that display most of the eastern provinces in the Low-Low quadrant and the western ones in the High-High quadrant. Finally, LISA statistics confirm the significant presence of local spatial autocorrelation and highlight spatial heterogeneity in the form of two distinct spatial clusters of high and low values of per capita income. Overall, these results confirm the dualistic structure of Turkey’s economic geography, as many previous studies had showed.

References


Anselin L (1995) Local Indicator of Spatial Association – LISA. Geographical Analysis. 27, 93–115


Tukey JW (1977) *Exploratory Data Analysis*. Addison-Wesley, Reading, MA